

Sujet de stage M2 « Analyse de données textuelles sur la sécurité alimentaire en Afrique de l'Ouest »

Le stage s'inscrit dans le cadre d'un projet pluridisciplinaire concernant la gestion des risques liés à la **sécurité alimentaire en Afrique de l'Ouest**, considérée comme l'un des enjeux majeurs de développement de la région. Parmi les raisons à l'origine de ce phénomène, nous pouvons citer une forte croissance démographique, une agriculture pluviale très dépendante des conditions pluviométriques, auxquels s'ajoutent des risques sécuritaires et sanitaires. Depuis les grandes sécheresses du début des années 70, plusieurs systèmes d'alerte précoce (SAP) de la sécurité alimentaire ont été développés sur la région pour permettre aux décideurs d'anticiper les crises, et d'aider à la planification des mesures d'urgence en ciblant les populations et/ou les zones à risques. Dans ces systèmes, l'information satellitaire est utilisée majoritairement pour dériver des anomalies d'indices de végétation à partir de séries temporelles d'images à basse résolution spatiale. Les organisations internationales en charge des différents systèmes de suivi et d'alerte, se réunissent mensuellement pour atteindre un consensus sur les conditions de la campagne agricole. Si les classifications sur l'état des cultures sont souvent cohérentes, il arrive que ces informations divergent ou soient en contradiction avec les observations de terrain (Becker-Reshef et al., 2020)¹. Ces désaccords peuvent venir des différences en termes de couverture géographique, d'unités spatiales cartographiées, de mandat des organisations en charge des SAPs, et des méthodes mises en œuvre. Dans ce contexte, les données textuelles (par exemple, articles de journaux) représentent une source d'information inexploitée, qui peut être utilisée pour renforcer les SAPs et résoudre les situations de désaccord.

L'objectif de ce stage est d'utiliser et combiner des techniques avancées de fouille de texte et de traitement automatique du langage naturel (TALN) à un corpus de données textuelles sur le thème de la sécurité alimentaire en Afrique de l'Ouest, afin d'apporter des informations complémentaires permettant de lever des incohérences observées et d'établir un diagnostic sur l'état de la végétation. Plus précisément, étant donné un cas d'étude spécifique (par exemple, pays et/ou épisode de désaccord), dans une première étape, des méthodes de l'état de l'art de *Topic Modeling* seront utilisées pour obtenir des sous-ensembles de données thématiquement homogènes. Le stage sera focalisé sur des documents textuels en Français, ce qui représente un autre défi scientifique vu la mineur quantité de ressources dans l'état de l'art par rapport à l'Anglais. Une fois ces clusters obtenus pour chaque cas d'étude, différentes approches pourront être testées pour la phase de recherche de consensus :

- Approches fondées sur des techniques de *Sentiment Analysis* et *Opinion Mining* afin de comparer les polarités d'opinion (positif, négatif, neutre) ;
- Approches supervisées fondées sur des techniques de Machine Learning. Dans ce cas, l'idée est d'exploiter des données labélisés pour entraîner un classificateur de textes, afin de reconnaître une situation favorable ou défavorable à l'état des cultures. Le classifieur sera ensuite utilisé pour classifier les sous-ensembles de documents textuels associé aux épisodes de désaccord.

Des modèles de langages pour la langue Française basés sur la technologie des Transformers (p.ex., CamemBERT, FlauBERT) pourront aussi être utilisés pour supporter les deux tâches. Ces méthodes devront

être combinées pour apporter des connaissances nouvelles. Dans ce travail, les différentes propositions devront intégrer les dimensions spatio-temporelles associées aux données textuelles qui devront être prises en compte dans les analyses réalisées. Ces dernières seront effectuées à partir de cas d'étude déterminés permettant d'évaluer les différentes propositions. Ainsi, le ou la stagiaire contribuera à la constitution d'un corpus de données textuelles sur la sécurité alimentaire en Afrique de l'Ouest liés aux cas d'étude. Enfin, le travail sera valorisé à travers l'écriture d'un article scientifique qui présentera les contributions méthodologiques et les résultats obtenus.

¹Becker-Reshef I., Justice C., Barker B., et al., 2020. Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. Remote Sensing of Environment, 237 : 111553. <https://doi.org/10.1016/j.rse.2019.111553>

Compétences requises :

Langage Python, outils NLP
Capacité de travail en équipe pluridisciplinaire.

Divers :

Durée : 6 mois
Gratification : taux légal en vigueur
Localisations : TETIS (Maison de la Télédétection) à Montpellier

Candidature :

Envoyer un CV + relevés de notes des deux dernières années à roberto.interdonato@cirad.fr et mathieu.roche@cirad.fr

UMR TETIS
500 rue Jean-François Breton
34000 Montpellier – France

Tél.: +33(0)4 67 54 87 54
<http://www.umr-tetis.fr>